

The Balanced Scorecard as a Strategy-Evaluation Tool: The Effects of Implementation Involvement and a Causal-Chain Focus

William B. Tayler
Emory University

ABSTRACT: Using an experiment, I examine whether involvement in scorecard implementation can mitigate the effects of motivated reasoning that occur when the scorecard is framed as a causal chain rather than merely as a balanced set of measures. Psychological research on motivated reasoning suggests that managers will evaluate and interpret data in ways consistent with preferences, increasing their tendency to arrive at conclusions that are consistent with their desired conclusions (Kunda 1990). Consistent with that research, results of my study show that managers who are involved in selecting strategic initiatives perceive those initiatives as having been more successful than managers who are not involved in the initiative-selection process (holding constant actual scorecard performance). Results suggest further that simply framing the scorecard as a causal chain is not sufficient to mitigate these effects. However, framing the scorecard as a causal chain, in conjunction with involving managers in the selection of scorecard measures, mitigates the effects of motivated reasoning tied to manager involvement in initiative selection.

Keywords: *balanced scorecard; motivated reasoning; causal chain; business strategy; measure selection; strategic performance measurement system.*

I am indebted to my dissertation chair, Rob Bloomfield, as well as to the members of my dissertation committee, Tom Gilovich, Bob Libby, and Mark Nelson, and to Marlys Lipe (editor). I am grateful for helpful comments from two anonymous reviewers for this journal, an anonymous reviewer from the AAA Annual Meeting, as well as from Sarah Bonner, Harry Evans, Paul Fischer, Sue Haka, Gary Hecht, Ron Hilton, Frank Hodge, Steve Kachelmeier, Kathryn Kadous, Jane Kennedy, Ron King, Susan Krische, Joan Luft, Laureen Maines, Don Moser, Grace Pownall, Steve Salterio, Lisa Sedor, Mike Shields, Steve Smith, Karen Sedatole, Geoff Sprinkle, Kristy Towry, Sandra Vera-Muñoz, Greg Waymire, and workshop participants at Brigham Young University, Cornell University, Emory University, Indiana University, The University of Illinois at Urbana-Champaign, The University of Iowa, Michigan State University, University of Notre Dame, The Pennsylvania State University, University of Pittsburgh, University of Southern California, The University of Utah, University of Washington, and Washington University in St. Louis. I gratefully acknowledge the financial support from the American Accounting Association and the Institute of Management Accountants Outstanding Doctoral Dissertation Award, Cornell University's Johnson Graduate School of Management, the Institute of Management Accountants Doctoral Student Grant Program, and Emory University's Goizueta Business School.

Editor's note: Accepted by Marlys Lipe.

Submitted: July 2007
Accepted: September 2009
Published Online: May 2010

I. INTRODUCTION

The balanced scorecard has been heralded as one of the most significant developments in management accounting (Atkinson et al. 1997).¹ This management tool has evolved considerably over the past two decades. Early writings on the balanced scorecard emphasized the importance of using multiple measures to provide a balanced perspective of firms' performance, and established one of the most salient features of balanced scorecards: the grouping of measures into four distinct categories of performance (financial, customer, internal processes, and learning and growth; Kaplan and Norton 1992). More recently, balanced-scorecard proponents have focused on the need to tie measures together into a causal chain of performance and to test the validity of these hypothesized effects to guide the development of strategy. Kaplan and Norton (2001) argue that one of the primary benefits of the balanced scorecard is its use in gauging the success of strategy.

Archival and field research has yielded mixed results on the benefits of balanced-scorecard usage for strategy-evaluation purposes. One potential factor limiting the benefits of scorecard usage in some settings is the propensity of decision-makers' reasoning to be influenced by their motivations. Research in psychology suggests that this "motivated reasoning" begins with the manner in which evidence is assembled and evaluated (Kunda 1990; Gilovich 1991; Ditto and Lopez 1992; Ditto et al. 1998; Dawson et al. 2002; Ditto et al. 2003; Dawson et al. 2006a). Settings in which evaluators have preferences to arrive at certain conclusions, and for which information is noisy, ambiguous, or complex, provide fertile ground for motivated-reasoning processes because evaluators can more easily rely primarily on preference-consistent information while maintaining an appearance of objectivity. A balanced-scorecard environment provides just such a setting, emphasizing the use of multiple (noisy) measures of performance for which evaluators often have an active role in the selection of strategy. Thus, the objective of this study is to investigate the effect of involvement in the selection of strategy on subsequent evaluation of that strategy using the balanced scorecard, and to investigate features of the scorecard implementation process that can mitigate the impact of strategy-selection involvement.

I use an experiment to examine whether managers involved in the selection of a strategic initiative will succumb to motivated reasoning when using a balanced scorecard to evaluate the success of the new initiative. I further investigate methods to help offset the impact of motivated reasoning on strategy evaluation. First, I examine whether simply framing the scorecard as a causal chain, rather than merely as a balanced set of measures, can help limit the effect of motivated reasoning on strategy evaluations. If the scorecard is framed as a causal chain, then managers should be less susceptible to motivated reasoning because success becomes more clearly defined: an initiative must not only improve the measure on which it should have a direct effect (such as customer satisfaction), but must also improve the measures on which it should have an indirect effect (such as sales). Following Kunda (1990), I expect that this more structured view will reduce managers' latitude to perceive unsuccessful initiatives they selected as successful.

Next, I investigate whether managerial involvement in measure selection increases the effectiveness with which a causal-chain framing of the scorecard offsets tendencies to assess outcomes too favorably. Managers involved in measure selection will likely be motivated to perceive a measure they selected as a good one (Kunda 1990). If the scorecard is framed as a causal chain, then managers will want to see that the measure they selected is a good predictor of performance further down the causal chain, increasing attention to these linkages. If managers cannot find

¹ A recent study by Bain & Company indicates that 53 percent of firms worldwide use the balanced scorecard, including 61 percent of large firms and 49 percent of firms in North America (Rigby and Bilodeau 2009).

sufficient evidence to support their preferred belief regarding the measure they were involved in selecting, then this will likely also cast doubt on the ability of the strategic initiative they selected to impact performance further down the causal chain.

In my experiment, M.B.A. students with a basic understanding of balanced-scorecard concepts play the role of managers who are evaluating a new strategic initiative at their firm. Participants are provided with balanced-scorecard data upon which to base their analyses, and answer questions regarding the success of the initiative and the process they used to perform their assessment. In all experimental settings, scorecard data indicate that the new initiative has a strong effect on customer satisfaction, but that neither initiative nor customer satisfaction has an effect on financial performance. This structure provides latitude for participants to form conclusions consistent with their preferences. Participants with motivation to perceive the strategic initiative as successful can focus on the strong effects of their initiative on customer satisfaction and deemphasize or ignore the lack of effects on financial performance.

The experimental design consists of two levels of scorecard framing crossed with three levels of scorecard implementation involvement (creating six between-subjects conditions). The scorecard is framed either in the traditional “four groups” format, which groups performance measures into the four classic scorecard perspectives (Kaplan and Norton 1992), or in a “causal chain” format, which emphasizes hypothesized causal linkages (Kaplan and Norton 1996b, 2000). The three levels of scorecard implementation involvement are cumulative in nature, with some participants making no decisions regarding strategic initiatives or scorecard measures, some selecting the strategic initiative to be pursued but not the scorecard measures, and some selecting both the initiative and a measure of performance tied to that initiative.

Consistent with prior research on motivated reasoning, results indicate that managers who are involved in initiative selection are more likely to recommend rolling out the initiative firm-wide, holding constant initiative performance. Results further suggest that merely framing the balanced scorecard as a causal-chain is not sufficient to diminish motivated-reasoning processes, even in this simple setting. However, framing the scorecard as a causal chain *and* involving managers in measure selection mitigates the effects of motivated reasoning related to managers’ involvement in initiative selection.

Debriefing data help to establish the process underlying these effects by providing insight into what information participants emphasized in their evaluations. Participants with greater motivation to perceive the initiative as successful (those who are involved in the selection of the initiative) place less emphasis on scorecard data indicating otherwise than participants who are not involved in initiative selection. This differential focus on preference-inconsistent information partially mediates the primary results.

These results contribute to research on the balanced scorecard in multiple ways. First, the study demonstrates the effects of a causal view of the balanced scorecard in a strategy-evaluation task. Proponents of the balanced scorecard emphasize the importance of tying performance measures to strategy and hypothesizing a testable causal chain of performance, in part to ensure that important indirect results follow the more direct successes (Kaplan and Norton 2001). However, empirical evidence suggests that firms often do not explicitly state causal chains (Ittner and Larcker 2003; Malina and Selto 2004). This study provides evidence that a causal view of the scorecard, in conjunction with involvement in scorecard measure selection, helps overcome psychological forces that are likely to limit scorecard effectiveness.

Second, this study provides insights on the balanced-scorecard implementation process. Scorecard proponents have begun to address issues that arise when trying to implement scorecards, such as which parties should be involved with the selection of scorecard measures (Frigo and Krumwiede 2000; Niven 2002; Kaplan and Norton 2006). Scorecard proponents often focus on managerial “buy-in,” suggesting that involvement in measure selection by those affected by the

scorecard will increase the impact of the scorecard (e.g., Cokins 2005). My study suggests that another important benefit derived from increased managerial involvement in the selection of scorecard measures is that it can counter motivated reasoning, thereby reducing the likelihood that managers will view unsuccessful strategies as successful.

Third, this study provides insights on the use of the scorecard for strategy evaluation. Scorecard proponents have emphasized that the scorecard is not only for performance evaluation, but is also a tool for developing and evaluating strategy (Kaplan and Norton 2000, 2001; Niven 2002; Buytendijk et al. 2004; Kaplan and Norton 2004a, 2004b, 2006). However, most prior balanced-scorecard research has focused solely on managers' use of the scorecard for performance evaluations. My study focuses on the scorecard as a tool for defining and refining strategy.

Finally, this study contributes to psychological research on motivated reasoning by demonstrating how the effects of motivated reasoning can be diminished through the introduction of additional motivations. This "fighting fire with fire" approach essentially accepts motivated reasoning as given, but adds motivations through adjusting decision-maker involvement in the task to help balance prior motivations to arrive at certain conclusions (see Bonner (2007, §10-1) for a discussion of the importance of investigating the effects of conflicting motivators in accounting).

Section II provides the background and hypotheses. Section III describes the experimental design and related procedures. Section IV discusses results. Section V summarizes results and discusses implications, limitations, and directions for future research.

II. BACKGROUND AND HYPOTHESES

The Balanced Scorecard

Early writings on the balanced scorecard focused on the ability of multiple measures to provide a more balanced perspective of firms' performance (Kaplan and Norton 1992). Under this view, the four scorecard categories (financial, customer, internal processes, and learning and growth) keep managers from focusing solely on financial-performance measures. The majority of early research on the balanced scorecard has focused on the "balance" of the scorecard, investigating how managers use scorecard measures to evaluate performance (Lipe and Salterio 2000, 2002; Ittner et al. 2003b; Banker et al. 2004; Libby et al. 2004; Roberts et al. 2004; Dilla and Steinbart 2005).

Recently, scorecard proponents have shifted emphasis from balance to strategy, arguing that the scorecard serves as a tool for defining strategic objectives and communicating them throughout the organization, identifying initiatives to achieve those objectives, and evaluating whether those objectives have been achieved (Kaplan and Norton 2000, 2001; Niven 2002; Buytendijk et al. 2004; Kaplan and Norton 2004a, 2004b, 2006).² Scorecards are tied to strategy through the "strategy map" (Kaplan and Norton 2000), also called a "value driver map" (Ittner and Larcker 2003). Strategy maps translate expected results into testable hypotheses to enhance "strategic learning," the process of using the strategically aligned scorecard measures as a way of measuring the success of strategy (Kaplan and Norton 2001). If linkages in the hypothesized causal chain of performance prove spurious, then the scorecard, or the strategy that drives it, can be adjusted.

In conjunction with scorecard proponents' shift in emphasis from balance to strategy, research on the balanced scorecard has begun to look at the use of the balanced scorecard in strategy development (Malina and Selto 2001; Ittner et al. 2003a; Ittner and Larcker 2003; Campbell et al. 2008). Malina and Selto (2001) provide field evidence indicating that scorecards do not always

² Scorecard objectives are statements that define the purpose of actions and form the basis of measures used in performance and strategy evaluation. Scorecard initiatives are actionable plans intended to affect performance in targeted objectives.

have explicitly defined causal linkages. [Ittner and Larcker \(2003\)](#) corroborate this finding with survey data showing that less than 30 percent of firms that use the balanced scorecard have explicitly stated causal chains, and that when strategy maps are used, managers often fail to test the hypothesized causal chain. In their study, only 21 percent of firms with explicitly stated causal chains actually test the validity of effects suggested by the causal chains. [Campbell et al. \(2008\)](#) provide field evidence from one firm that appears to have successfully tested the hypothesized causal chain and adjusted its strategy accordingly.

Balanced-scorecard implementation issues have also received increased emphasis in recent years (e.g., [Niven 2002](#); [Kaplan and Norton 2006](#)). Viewed narrowly, scorecard implementation involves (among other things) the selection of measures, the collection of scorecard-related data, the formatting of scorecard reports, and the dissemination of scorecard information. When the scorecard is viewed as a tool for defining, executing, and measuring strategy, scorecard implementation also involves the allocation of decision rights regarding strategy selection and plans for achieving strategic objectives. [Kaplan and Norton \(1996a\)](#) recommend that scorecard development be a joint effort of unit managers and upper management. [Cokins \(2005, 2\)](#) suggests that manager involvement in scorecard implementation generates “buy-in and ownership of the scorecard and key performance indicators,” which will increase the impact of the scorecard on the organization.

Though scorecard proponents have begun to address scorecard implementation, little academic research has been done on balanced-scorecard implementation issues. However, limited empirical evidence indicates that manager involvement in scorecard implementation varies between firms. [Kaplan and Norton \(2001\)](#) cite evidence of heavy involvement in initiative and measure selection on the part of unit managers, but [Malina and Selto \(2001\)](#) provide evidence suggesting that some firms impose scorecards on units without seeking input from those affected. [Lipe and Salterio \(2000\)](#) point out that scorecard-related judgments could be influenced by evaluator involvement in the implementation process.

Motivated Reasoning

Prior accounting research has demonstrated that judgments tend to be influenced by motivations to arrive at particular conclusions. [Cuccia et al. \(1995\)](#), [Cloyd and Spilker \(1999\)](#), and [Kadous et al. \(2008\)](#) provide evidence that tax professionals’ evaluation of evidence and information search are driven in part by their clients’ preferred conclusions. Other studies have focused on the effect of client preferences on auditors, demonstrating the biasing effects of auditors’ motivations on their judgments ([Hackenbrack and Nelson 1996](#); [Bazerman et al. 1997](#); [Bazerman et al. 2002](#); [Phillips 2002](#); [Beeler and Hunton 2002](#); [Wilks 2002](#); [Kadous et al. 2003](#); [Han et al. 2006](#); [Moore et al. 2006](#); [Nelson 2006](#)). These studies suggest the motivation to arrive at clients’ preferred conclusions is one of the reasons “why good accountants do bad audits” ([Bazerman et al. 2002, 96](#)). Additional studies have demonstrated this bias also persists among analysts ([Hunton and McEwen 1997](#)), investors ([Hales 2007](#); [Seybert and Bloomfield 2009](#); [Thayer 2009](#)) and management accountants ([Boiney et al. 1997](#); [Bloomfield and Luft 2006](#)).

Research in psychology has helped clarify the process behind the effects of motivations on judgments, showing that individuals tend to evaluate and interpret data in ways consistent with their preferences. This pervasive tendency has come to be known as “motivated reasoning” ([Kunda 1990](#)).

[Gilovich \(1991\)](#) proposes that motivated reasoning is driven by the way in which evaluators approach evidence. When encountering a disagreeable proposition, people tend to ask “Must I believe this?” and pursue information in an attempt to disconfirm or cast doubt on the validity of the bad news ([Ditto and Lopez 1992](#); [Ditto et al. 1998](#); [Dawson et al. 2002](#); [Ditto et al. 2003](#);

Dawson et al. 2006a). This high standard for acceptance leads to a more thorough information search and a stringent interpretation of evidential weaknesses, thus increasing the likelihood that an individual will find and accept evidence disconfirming the bad news. However, when encountering an agreeable proposition, people ask “Can I believe this?” and pursue information in an attempt to validate the good news. This more permissive acceptance standard allows for a more superficial evaluation of evidence (Ditto et al. 1998; Dawson et al. 2006a; Dawson et al. 2006b), biased compilation of data (Lord et al. 1979; Dunning et al. 1989), and earlier truncation of information search (Ditto and Lopez 1992; Ditto et al. 1998; Ditto et al. 2003; Dawson et al. 2006a). However, the acceptance of preference-consistent information is constrained by the reasonableness of the data (Kunda 1990). Even when succumbing to motivated reasoning, decision-makers attempt to maintain an “illusion of objectivity” (Pyszczynski and Greenberg 1987, 302) by “attempt[ing] to be rational and to construct a justification of their desired conclusion that would persuade a dispassionate observer” (Kunda 1990, 482–482).

A number of institutional features of the balanced scorecard produce fertile ground for motivated-reasoning processes. A common criticism of the balanced scorecard revolves around its emphasis on nonfinancial performance measures, which are typically more noisy, subjective, and ambiguous than traditional financial measures (Ittner and Larcker 2003). This decrease in precision and clarity provides managers with additional latitude in interpreting evidence. Further, even the most basic of performance-measurement systems provides multiple forms of performance feedback. The balanced scorecard is no exception. A manager in search of preference-consistent information has multiple potential sources of agreeable data. Once a reasonable amount of supportive data is accumulated, managers are likely to stop searching, disregard disconfirming evidence, or completely overlook or reinterpret conflicting data.

Managers who use balanced-scorecard feedback to evaluate a strategic initiative they selected would prefer to discover that their selected initiative is successful. Holding constant actual scorecard performance, if scorecard evidence is ambiguous or sufficiently complex or noisy to allow different assessments of the success of the strategic initiative, managers with different preferences will likely reach different conclusions. Thus, I predict that managers who are involved in the selection of a strategic initiative interpret balanced-scorecard data as indicating that the initiative is more successful than do managers who are not involved in the selection of the initiative.

H1: Managers using a balanced scorecard to analyze the success of a questionable initiative will perceive the initiative as more successful if they were involved in the selection of the initiative.

Causal Chain

One potential method for reducing motivated reasoning among balanced-scorecard evaluators is to increase emphasis on the hypothesized causal chain. Framing the scorecard as a causal chain should decrease motivated reasoning by focusing managers’ attention on the need to test hypothesized linkages between scorecard components in making an assessment of the success of a strategic initiative, thereby reducing the latitude managers have to reasonably perceive data in a manner consistent with their preferences (Kunda 1990). Additionally, providing managers with a predicted causal model is likely to reduce the cognitive complexity of evaluation tasks where multiple measures of performance are available for use (Morecroft et al. 2002; Malina and Selto 2004; Krishnan et al. 2005). In turn, reduced complexity is likely to reduce ambiguity in feedback

and restrict managers' ability to see only what they want to see in the data.³ Further, when managers understand all of the anticipated cause-and-effect linkages that follow from a strategic initiative, they should be less persuaded by individual instances of apparent success (e.g., an increase in customer satisfaction) when other hypothesized effects do not follow (e.g., an increase in financial performance). Because decision makers who succumb to motivated reasoning are constrained in their acceptance of good news (and their rejection of bad news) by their ability to maintain an "illusion of objectivity" (Pyszczynski and Greenberg 1987; Kunda 1990), a clearly defined causal chain should limit evaluators' ability to arrive at a preference-consistent conclusion if linkages deriving from an initiative they were involved in selecting prove inaccurate.

H2a: Managers involved in initiative selection will be less likely to perceive a questionable initiative as successful if the scorecard is framed as a causal chain of performance.

Involvement in Measure Selection

A major aspect of the balanced-scorecard implementation process is the selection of performance measures. Niven (2002, 146) calls scorecard measures the "centerpiece of the scorecard system." Frigo and Krumwiede (2000, 51–52) emphasize the importance of involving middle management in measure selection, noting that "departments know their key performance measures and key success factors better than anyone else ... and, therefore, they are in the best position to develop their own scorecards."

Just as managers who are involved in the selection of an initiative are motivated to perceive the initiative as successful, managers who are involved in measure selection are motivated to perceive that their choice of a performance measure is a good one. With a causal-chain framing, one important feature of a good measure is that it can be used as a predictor of performance further down the causal chain. Thus, involvement in measure selection should increase attention to these causal linkages. For example, if a manager who believes that customer satisfaction is causally linked to financial performance elects to monitor customer satisfaction through a measure of customer retention, then that manager would prefer to discover that increased customer retention leads to improvements in financial performance and will look for this relation. Thus, the motivated reasoning generated by manager involvement in *measure selection* has the potential to mitigate the effects of managers' motivated reasoning tied to their involvement in *initiative selection* by influencing the body of evidence evaluated by managers. Many initiatives have strong direct effects on performance but fail to drive performance further down the causal chain. For example, an initiative can improve customer satisfaction (a direct effect) without increasing financial performance (the ultimate, indirect goal). If managers in this setting are motivated to perceive an initiative they selected as successful, but are also motivated to find that the performance they opted to measure drives financial performance (for example), then their evaluation of the evidence is more likely to uncover the (disappointing) disconnect in the causal chain.

Managers succumbing to motivated reasoning are constrained in their ability to arrive at preference-consistent conclusions by the availability of sufficient evidence to reasonably support such an inference (Pyszczynski and Greenberg 1987; Kunda 1990). Thus, the inability to support a preferred belief of a causal link between the selected measure and performance further down the causal chain will cast doubt not only on the quality of the measure, but also on the viability of an initiative that was expected to affect performance further down the causal chain. Importantly, this

³ Relatedly, Farrell et al. (2009) show that providing the predicted causal model to decision makers whose actions affect performance reduces the tendency to choose actions that maximize current period payoff at the expense of overall payoff (Herrnstein 1990; Herrnstein and Prelec 1991, 1992; Mainwaring 1997). Also, Vera-Muñoz et al. (2007) demonstrate that provision of a predicted causal model enhances managers' ability to exploit covariation information in data.

reasoning holds only if managers have an understanding of the causal chain *and* managers are motivated to evaluate evidence in light of this tool because of their involvement in measures selection.

H2b: Managers involved in initiative selection will be less likely to perceive a questionable initiative as successful if the scorecard is framed as a causal chain *and* managers are involved in measure selection.

III. METHOD

Task and Design

Participants assumed the role of managers over Paladin Pizza, a fictitious pizza chain (hereafter referred to as the “firm”), in an experimental task involving the use of balanced-scorecard data in order to evaluate a new strategic initiative to determine whether that initiative should be rolled out firm-wide. The experimental materials provided all participants with background information for the firm and indicated that the firm was considering two different initiatives for implementation: (1) the “side order strategy,” in which stores give a free side order for every five pizzas purchased, and (2) the “quality ingredients strategy,” in which stores use high-quality ingredients (relative to current quality) for pizzas.⁴ Participants were also told that the firm was considering two different measures relating to the new initiative: (1) the “customer survey score,” which is based on a customer survey measure of customer intentions to return to order pizza, and (2) the “returning customer score,” which is based on the actual number of return customers (based on credit card data).

The experiment crossed two types of scorecard framing with three levels of scorecard implementation involvement. To manipulate scorecard framing, participants in the “four groups” (FG) scorecard-framing setting received background information that described the scorecard as being composed of four categories of performance (financial, customer, internal processes, and learning and growth), whereas participants in the “causal chain” (CC) scorecard-framing setting received background information that expanded on this description, emphasizing hypothesized cause-and-effect relationships between scorecard components. To reinforce the written description of the scorecard-framing manipulation, background information also provided a graphical depiction of the four groups or casual chain similar to that in [Banker et al. \(2004\)](#), as shown in Panels A and B of Figure 1.

I manipulate scorecard implementation involvement at three levels. Participants in the “low involvement” condition received case materials that presented the two initiatives under consideration, followed by a sentence indicating that top management had decided to pursue the “quality ingredients strategy.” Case materials then presented participants with the two measures under consideration, followed by a sentence indicating that the accounting department had decided to use the “returning customer score.” This “low involvement” condition provides a baseline setting for which participants make judgments absent involvement in scorecard implementation. Thus, judgments in this setting are not affected by motivated reasoning tied to involvement in the selection of the strategic initiatives or scorecard measures.

Participants in the “initiative selection involvement” condition received identical information as those in the “low involvement” condition, except that following the presentation of the two initiatives under consideration, case materials indicated that the final decision regarding which

⁴ Case materials indicated that both initiatives would generate equivalent increases in variable costs, neither initiative would create additional fixed costs, and prices would not be increased in conjunction with either of the initiatives.

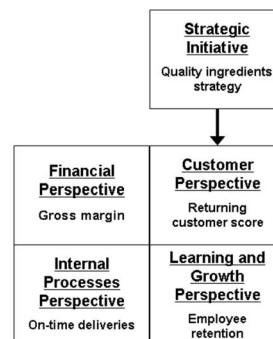
FIGURE 1
Scorecard Framing

Half of the participants received background information describing the scorecard as being composed of four categories of performance (financial, customer, internal processes, and learning and growth), as shown in Panel A below. The other half of the participants received the first paragraph in Panel A, as well as additional information emphasizing hypothesized cause-and-effect relationships between categories, as shown in Panel B below. To reinforce the written description of the scorecard-framing manipulation, participants were also presented with a graphical depiction of the balanced scorecard.

Panel A: Four Groups Framing

Paladin Pizza uses a “balanced scorecard” to measure performance at each of its restaurants. The balanced scorecard reflects the reality that strategic success depends not only on strong financial performance, but also on performance in a variety of other dimensions. In particular, a typical scorecard emphasizes the following four categories: financial, customer, internal processes, and learning and growth. The balanced scorecard is not just a measurement tool. Companies use their scorecards to focus on improvement in areas thought to be particularly important to strategic success.

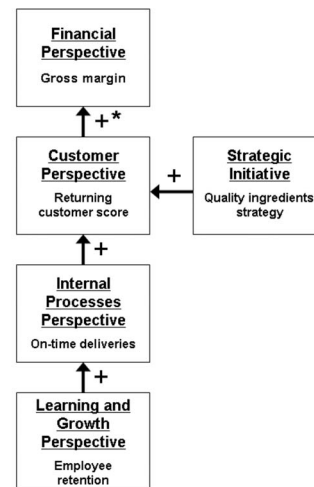
Please take a moment to examine the figure to the right, which depicts the four scorecard dimensions at Paladin Pizza, the measures in those dimensions, and the new strategic initiative being pursued.



Panel B: Causal Chain Framing

The balanced scorecard reflects a step-by-step sequence of cause-and-effect relationships leading from the most fundamental aspects of performance to financial performance. In particular, the typical scorecard identifies hypothesized causal links between scorecard categories. Strong learning and growth leads to improved internal processes. Improved internal processes lead to increased customer satisfaction. Increases in customer satisfaction leads to improved financial performance.

Please take a moment to examine the figure to the right, which depicts the hypothesized linkages between the four scorecard dimensions at Paladin Pizza, the measures in those dimensions, and the new strategic initiative being pursued.



* "+" above indicates a positive hypothesized correlation

strategic initiative to pursue was theirs. Participants were then asked to select which of the strategies they believed the firm should pursue by placing a checkmark next to their choice and providing a brief explanation for their selection.

Participants in the “initiative and measure selection involvement” condition received identical information as the “initiative selection involvement” participants, except that following the presentation of the two measures under consideration, case materials indicated that the final

decision regarding which measure to implement was theirs. Participants were then asked to select which of the measures they believed the firm should use by placing a checkmark next to their choice and providing a brief explanation for their selection.

Procedure

Participants began the session by filling out an informed-consent form. They then read brief instructions and were directed to open an envelope containing case materials. Participants then read background information for the firm and information regarding the initiatives and measures under consideration (with some participants selecting their preferred initiative and/or measure, depending on the scorecard-implementation-involvement condition to which they had been randomly assigned).

Next, case materials directed participants to open another envelope containing background information on the balanced scorecard for the firm. Case materials directed participants to one envelope in the “low involvement” condition, but directed participants in the “initiative selection involvement” condition to open one of two envelopes (based on their initiative selection), and directed participants in the “initiative and measure selection involvement” condition to open one of four envelopes (based on their initiative and measure selections), so that the scorecard-component labels reflected participants’ actual scorecard-component choices. However, the background information contained in all envelopes and scorecard-implementation-involvement conditions was held constant, varying only in the labels applied to the scorecard components to be consistent with the initiative or measure selected (e.g., a participant who selected the “returning customer score” would see this label reflected on the scorecard, rather than “customer survey score”). For simplicity, only the customer perspective had a stated objective (“delight the customer”) beyond that implied by the perspective title and associated measures. Also, for simplicity, each perspective had only one measure of performance.⁵

After reading background information regarding the firm’s scorecard, participants received balanced-scorecard data from 14 restaurants. To avoid a feedback confound, scorecard data were identical regardless of experimental condition, initiative choice, or measure choice (aside from the labels applied to the specific scorecard components).⁶ Scorecard data for an entire calendar year following initiative implementation were presented, along with results as a percentage of the previous year’s results. Participants were told that the new initiative was implemented on a trial basis at only 7 of the 14 restaurants and that “strategy consultants have indicated that any effects of the new strategy will be visible well within the first year of implementation.”⁷

Figure 2 presents the balanced-scorecard data given to all participants. The data show that the initiative improved the customer-perspective measure (either the “returning customer score” or the “customer survey score,” depending on treatment and choice of measure; $t = 4.69$, $p < 0.001$), but

⁵ Many scorecards have multiple objectives in each of the scorecard perspectives, and each objective can have multiple measures of performance. Further, firms can opt to divide scorecards into more (or less) than four perspectives—deviating from the classic format originally prescribed by Kaplan and Norton (1992), and can provide performance targets or benchmarks for some (or all) of the measures. Though technical attributes of the scorecard may affect the levels of the effects examined in this study, they are unlikely to affect the directional results (Libby et al. 2002).

⁶ Realizations of both the “customer survey score” and the “returning customer score” use the same scale, allowing all numeric data to be held constant, regardless of initiative and measure selections.

⁷ Some participants may have believed that effects of the initiative on financial performance would take more than one year to materialize. However, idiosyncratic beliefs of this nature do not explain differences between cells, as subjects were randomly assigned to treatment groups and scorecard data and background information was held constant between cells.

FIGURE 2
Scorecard Data

Participants read the following regarding the data below (the wording varied slightly based on scorecard-implementation-involvement condition and the actual initiative and measure selected): "Below is balanced-scorecard data from the 14 Paladin Pizza restaurants for the calendar years 2004 and 2005. The new measure you selected (the 'Returning customer score') was implemented in January 2004, and the new strategic initiative you selected (the 'Quality ingredients strategy') was implemented in January 2005. Note that, as indicated on the scorecard below, the strategy was implemented on a trial basis at only 7 of the 14 restaurants. Strategy consultants have indicated that any effects of the new strategy will be visible well within the first year of implementation. For each measure, average performance for 2005 is given, as is 2005 performance as a percent of 2004 performance."

RESTAURANT #:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Financial Perspective														
<i>Gross margin (in millions)</i>														
2005 results	3.65	4.01	3.02	3.72	3.68	3.33	2.79	3.16	3.25	4.19	2.90	3.69	3.26	3.92
2005 as % of 2004	103%	101%	98%	101%	100%	102%	95%	100%	97%	110%	100%	107%	104%	108%
Customer Perspective														
<i>Returning customer score</i>														
2005 results	69.48	68.15	65.06	67.94	73.67	70.60	70.55	59.81	58.96	63.03	67.80	61.14	55.99	54.75
2005 as % of 2004	109%	112%	114%	108%	120%	113%	110%	88%	100%	100%	114%	100%	96%	103%
"Quality ingredients strategy" implemented?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No
Internal Processes Perspective														
<i>On-time orders</i>														
2005 results	0.83	0.80	0.79	0.78	0.81	0.80	0.76	0.81	0.78	0.82	0.85	0.78	0.79	0.81
2005 as % of 2004	109%	105%	101%	102%	105%	105%	100%	104%	100%	102%	110%	107%	105%	103%
Learning and Growth Perspective														
<i>Employee retention</i>														
2005 results	11.43	11.48	11.85	12.20	12.38	13.75	12.30	13.02	11.45	12.24	12.07	12.35	12.47	11.70
2005 as % of 2004	100%	97%	113%	109%	112%	120%	110%	115%	113%	115%	112%	112%	100%	96%

not the "financial-perspective" measure (gross margin for all conditions; $t = -0.11$, $p = 0.917$).⁸ There is also no relation between the customer-perspective measure and the financial-perspective measure ($r = -0.098$, $p = 0.739$).⁹ Thus, participants motivated to perceive the new initiative as successful (i.e., participants who were involved in the selection of the initiative) would be able to collect evidence in favor of their preference for a strong causal link between implementation and the measure of customer satisfaction, but data were also present that should cast doubt on the preferred conclusion (e.g., the lack of a relation between initiative implementation or the measure of customer satisfaction and gross margin).

After reviewing the balanced-scorecard data, participants rated the success of the new initiative on a scale of 0 to 100. Because no definition of "success" was provided in the study, participants were also asked the less-ambiguous question of how likely they would be to recommend rolling out the new initiative in the remainder of the restaurants (on a scale of 0 to 100).¹⁰ Participants then responded to debriefing and demographic questions.

⁸ Though implementation of the new initiative has a nonsignificant negative effect on raw gross margin, implementation has a marginally significant negative effect on gross margin as a percent of the previous year ($t = -1.79$, $p = 0.098$), casting further doubt on the success of the initiative.

⁹ A spurious linkage between customer satisfaction and financial performance is not uncommon in practice. Prior research has found evidence that both supports and casts doubt on the relation between these two perspectives (e.g., see American Quality Foundation (1992); Anderson et al. (1994); Banker et al. (2000); Smith and Wright (2004)).

¹⁰ Consistent with pilot study results, the majority of participants (74 percent) selected the "quality ingredients" initiative. For cells in which participants could select the initiative to be implemented, judgments regarding the propensity to roll out the initiative firm-wide were not affected by which initiative participants selected ($F = 1.03$, $p = 0.313$), nor were judgments regarding the success of the initiative ($F = 0.74$, $p = 0.393$). Also consistent with pilot study results, the majority of participants (70 percent) selected the "returning customer" measure. For cells in which participants could

Participants

Students in an M.B.A.-level management accounting course at a *BusinessWeek* top-20 M.B.A. program participated in the study as part of an in-class exercise. Prior to data collection, participants had all participated in class sessions dealing with the use of management-accounting data in decision making, and had been exposed to the balanced scorecard and some of its basic features (e.g., the division of performance measures into four perspectives). Participants were not taught in class about cause-and-effect relations among scorecard components until after they had taken part in the experiment. Thus, participants represented an ideal subject pool for the current study due to their limited exposure to some of the specific features of the balanced scorecard (specifically, strategy maps), the provision of which was manipulated in the experiment. This approach follows a tradition of using classroom instruction to control for knowledge (Butt 1988; Nelson 1993; Nelson et al. 1995; Bonner et al. 1996; Bonner et al. 1997; Nelson and Tayler 2007).¹¹

In all, 135 participants provided data during three successive sections of the same accounting course. Three participants opened the wrong envelope (based on the initiative and/or measure selection they checked), meaning that these participants received data for a different initiative or measure than the one they selected, potentially contaminating the scorecard-implementation-involvement manipulation. Thus, data from these participants are omitted from analyses, leaving 132 independent observations.¹² On average, participants had 5.3 years of work experience and were 28.4 years old. Sixty percent of the participants were male.

IV. RESULTS

Panel A of Table 1 shows the mean and standard deviation of rollout judgments by experimental cell; and Panel B provides a plot of the means. Panel C of Table 1 provides hypothesis test results.

Hypothesis 1 predicts that managers who select a strategic initiative are more likely to roll out the new initiative firm-wide than managers who did not select the initiative. Because H2 predicts that causal-chain framing will work against a main effect for initiative selection involvement, I first test H1 by comparing rollout judgments only between the “low involvement, four groups” condition and the “initiative selection involvement, four groups” condition. As shown in Panel C of Table 1, results indicate that the planned comparison is significant ($t = 2.24$, $p = 0.013$), with participants who are involved in initiative selection providing a mean rollout judgment (68.1) that is 16.1 higher than the mean rollout judgment of participants who are not involved in initiative selection (52.0).^{13,14} These results support H1.

Hypothesis 2a predicts that framing the scorecard as a causal chain will mitigate managers’ propensity to roll out the new initiative when they are involved in initiative selection. To test H2a,

select the measure to be implemented, subsequent judgments were not affected by which measure participants selected (rollout: $F = 0.07$, $p = 0.794$; success: $F = 0.56$, $p = 0.457$), nor by the combination (interaction) of their specific measure and initiative selection (rollout: $F = 0.15$, $p = 0.698$; success: $F = 1.85$, $p = 0.182$).

¹¹ Debriefing questions asked participants if they had previous experience with the balanced scorecard. Fifteen participants acknowledged having prior scorecard experience. Inferences are unchanged if data from these participants are omitted.

¹² Inferences are unchanged if data from these participants are included in the analyses.

¹³ p-values presented are one-sided when directional effects are hypothesized.

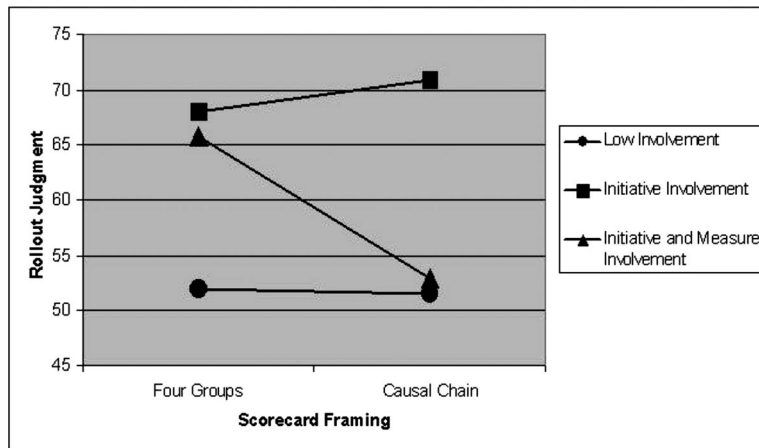
¹⁴ In addition to the rollout decision, participants were asked to rate the success of the new initiative. Rollout and success judgments were significantly, positively correlated ($r = 0.606$, $p < 0.001$). Similar to the rollout results, participants rated the initiative as more successful if they selected the initiative ($t = 1.63$, $p = 0.053$); however, no other effects based on this question are significant. Interpreting responses to this question is difficult, because the definition of what constitutes “success” of the initiative is not provided to participants, whereas the rollout question is less ambiguous. For example, a manager might rate the initiative as successful based on its effect on customer satisfaction, but choose not to roll out the strategy because of the lack of the expected effects on financial performance. For this reason, analyses and interpretation of results focus on the rollout decision, and the success rating is not discussed further.

TABLE 1
Rollout^a Judgment Means

Panel A: Mean Rollout Judgment (Standard Deviation)

Scorecard Framing ^b	Scorecard Implementation Involvement ^c			All
	Low Involvement	Initiative Selection Involvement	Initiative and Measure Selection Involvement	
Four Groups	n = 21 52.0 (20.9)	n = 21 68.1 (21.5)	n = 22 65.7 (23.8)	n = 64 62.0 (22.9)
Causal Chain	n = 25 51.5 (26.8)	n = 22 70.9 (22.2)	n = 21 52.9 (23.0)	n = 68 58.2 (25.5)
All	n = 46 51.7 (24.0)	n = 43 69.5 (21.6)	n = 43 59.4 (24.0)	n = 132 60.0 (24.2)

Panel B: Plot of Mean Rollout Judgments



(continued on next page)

Panel C: Hypothesis Tests

Planned Comparisons		Estimate	t- statistic	p-value (one-sided)
H1	Four Groups/Low Involvement versus Four Groups/Initiative Selection Involvement	16.1	2.24	0.013
H2a	Four Groups/Initiative Selection Involvement versus Causal Chain/Initiative Selection Involvement	2.8	0.39	0.695
H2b	Four Groups/Initiative Selection Involvement versus Causal Chain/Initiative and Measure Involvement	−15.2	−2.12	0.018

^a Participants rated how likely they would be to recommend rolling out the new initiative to the remainder of the Paladin Pizza restaurants, on a scale of 0 to 100. On the scale, 0 was described as “Very unlikely to recommend further implementation,” and 100 as “Very likely to recommend further implementation.”

^b Participants in the “four groups” scorecard-framing condition received background information describing the scorecard as being composed of four categories of performance (financial, customer, internal processes, and learning and growth), and were provided a simple diagram depicting these four categories. Participants in the “causal chain” scorecard-framing condition received background information that expanded on this description, emphasizing hypothesized cause-and-effect relationships between categories, and were provided a diagram depicting the categories and the causal chain they form.

^c Participants in the “low involvement” condition received case materials that described a new initiative and a new measure of customer satisfaction that would be implemented. Participants in the “initiative selection involvement” condition were asked to select a new initiative to implement, and were informed of a measure of customer satisfaction that would be implemented. Participants in the “initiative and measure selection involvement” condition were asked to select a new initiative and a measure of customer satisfaction for implementation.

I compare rollout judgments of participants in the “initiative selection involvement” cells. As previously noted, when the scorecard is framed as four groups, the mean rollout judgment is 68.1 when participants are involved in initiative selection. When the scorecard is framed as a causal chain, the mean rollout judgment is 70.9 when participants are involved in initiative selection. The difference of 2.8 is not significant ($t = 0.39$, $p = 0.695$). These results do not support H2, but instead suggest that simply framing the balanced scorecard as a causal-chain is not sufficient to overcome the effects of motivated reasoning for participants who are involved in initiative selection.

Hypothesis 2b predicts that managers’ propensity to roll out the new initiative decreases when managers are involved in measure selection and the scorecard is framed as a causal chain (holding constant that participants are involved in initiative selection). To test H2b, I compare rollout judgments in the “four groups, initiative involvement” cell to the “causal chain, initiative and measure involvement” cell. As noted above, when the scorecard is framed as four groups, the mean rollout judgment is 68.1 when participants are involved in initiative selection. When the scorecard is framed as a causal chain and participants are involved in both initiative and measure selection, the mean rollout judgment is 52.9. The difference of 15.2 is significant ($t = -2.12$, $p = 0.018$). Importantly, the significant difference between the “four groups, initiative involvement” cell to the “causal chain, initiative and measure involvement” cell appears to be driven by the combination of a causal-chain framing of the scorecard with manager involvement in measures selection, and not by involvement in measure selection alone. When the scorecard is framed as four groups and participants are involved in both initiative and measure selection, the mean rollout judgment is 65.8, which is not significantly different from rollout judgments when the scorecard is framed as four groups and participants are only involved in initiative selection (difference = 2.4, $t = -0.34$, $p = 0.737$). These results support H2b: managers who are involved in initiative selection are less likely to roll out the initiative when the scorecard is framed as a causal chain *and* managers are involved in choosing the measure of that initiative.

Interestingly, rollout judgments of participants in the causal-chain setting who were involved in both initiative and measure selection are not statistically different from those of participants in the “low involvement” conditions ($t = 0.19$, $p = 0.853$), suggesting that measure selection involvement coupled with a causal-chain framing is sufficient to completely mitigate the effects of motivated reasoning stemming from managers’ selection of an initiative in this setting.

Additional Analyses

Mediation Analyses

Prior literature suggests that motivated reasoning leads individuals to stop their search once they have collected sufficient preference-consistent data (e.g., that the new initiative increased customer satisfaction), or to perform only a superficial analysis or a biased compilation or interpretation of additional data (Lord et al. 1979; Dunning et al. 1989; Ditto and Lopez 1992; Ditto et al. 1998; Ditto et al. 2003; Dawson et al. 2006a; Dawson et al. 2006b). Motivated reasoning could take many forms in the current setting. For example, because financial performance did not improve with implementation of the new initiative (constituting bad news for those motivated to perceive the initiative as successful), participants who are involved in initiative selection will likely place less emphasis on the financial perspective. To test this conjecture, participants were asked the following: “Please rate how important each of the following perspectives on the balanced scorecard was in your assessment of the success of the strategy at Paladin Pizza by allocating 100 points among the four perspectives, allocating more points to perspectives that were

more important in your assessment.” The names of the four scorecard perspectives were then provided with a blank next to each perspective for participants’ point allocations.¹⁵

Results using financial-perspective emphasis as the dependent variable mirror those using rollout as the dependent variable. Consistent with motivated-reasoning processes, when the scorecard is framed as four groups, participants who are involved in initiative selection place less emphasis on the financial perspective than do participants in the “low involvement” condition. The mean financial-perspective emphasis is 44.6 when participants have low involvement and 32.3 when participants are involved in initiative selection, a difference of -12.3 ($t = -2.64$, $p = 0.005$). Further paralleling rollout-judgment results, there is no difference between financial-perspective emphasis of participants in the “initiative selection involvement” cells based on the framing of the scorecard ($t = -0.60$, $p = 0.546$). Also similar to rollout-judgment results, a comparison between the financial-perspective emphasis in the “four groups, initiative involvement” cell and the “causal chain, initiative and measure involvement” cell is marginally significant ($t = 1.29$, $p = 0.095$).

These results suggest a possible mediation of financial-perspective emphasis on the rollout judgment. Mediation requires that the following conditions hold: (1) a significant relation between the independent variable of interest (e.g., initiative involvement) and the dependent variable (e.g., the rollout judgment), (2) a significant relation between the independent variable (e.g., initiative involvement) and the potential mediator (e.g., financial-perspective emphasis), (3) a significant relation between the potential mediator and the dependent variable when controlling for the effect of the independent variable, and (4) a decline in the significance of the relation between the independent variable and the dependent variable when controlling for the effect of the potential mediator (James and Brett 1984; Baron and Kenny 1986; James et al. 2006).

Panels A and B of Table 2 present evidence that the effect of initiative selection involvement on the rollout judgment is mediated by financial-perspective emphasis. Step 1 shows the significant effect of initiative selection involvement in the four-groups setting ($t = 2.22$, $p = 0.014$).¹⁶ Step 2 shows the significant effect of initiative selection involvement on financial-perspective emphasis ($t = -2.64$, $p = 0.005$). Step 3 shows the significant effect of financial-perspective emphasis on the rollout judgment, when controlling for initiative selection involvement ($t = -2.39$, $p = 0.009$). Step 4 shows that the effect of initiative selection involvement on the rollout judgment is only marginally significant when financial-perspective emphasis is accounted for ($t = 1.64$, $p = 0.052$). The decrease in significance of the independent variable is significant (Freedman and Schatzkin (1992) $t = 1.79$, $p = 0.040$).¹⁷ Thus, the effect of initiative selection involvement on the rollout judgment works partially through the extent to which initiative selection involvement leads participants to decrease their emphasis on financial performance in their evaluations.¹⁸

¹⁵ The point allocations from four participants did not sum to exactly 100. I normalized allocations for these participants by dividing 100 by the point allocation sum and multiplying by the actual allocations. All analyses use the normalized allocations; however, inferences are unchanged using raw point allocations.

¹⁶ Seven participants did not answer the debriefing question regarding their balanced-scorecard perspective emphasis, and therefore should not be included in any portion of the mediation analyses. Dropping data from these participants causes a negligible change in the significance of the effect of initiative selection involvement ($p = 0.014$, rather than $p = 0.013$ as reported on Table 2).

¹⁷ See MacKinnon et al. (2002) for a review of mediation analysis methods, including the Freedman and Schatzkin (1992) test for the significance of mediation implied by a decrease in the significance of an independent variable when controlling for a potential mediator.

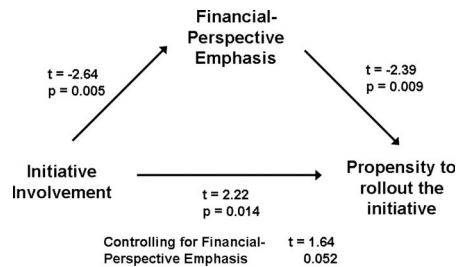
¹⁸ Results from a post-experimental survey are consistent with these results. One week following the experiment, participants were asked to respond to a brief survey to assess their recollection of financial performance data. Participants who were involved in initiative selection were more likely to be incorrect in their recall regarding financial performance data, suggesting that they were less likely to acquire and rely upon preference-inconsistent information ($t = 2.04$, $p = 0.022$).

TABLE 2
Financial-Perspective Emphasis Mediation Results

Panel A: Mediation Analysis Steps

	Mediation Analysis Steps			
	Step 1	Step 2	Step 3	Step 4
	Does Initiative Involvement ^a Predict the Rollout Judgment? ^b	Does Initiative Involvement Predict Financial- Perspective Emphasis? ^c	Does Financial- Perspective Emphasis Predict the Rollout Judgment When Controlling for Initiative Involvement?	Does Initiative Involvement's Significance Decrease (Relative to Step 1) When Controlling for Financial- Perspective Emphasis?
Estimate	16.0	-12.4	-0.3	11.9
t-statistic	(2.22)	(-2.64)	(-2.39)	(1.64)
p-value	[0.014]	[0.005]	[0.009]	[0.052] ^d

Panel B: Mediation Analysis Diagram



^a See Table 1, notes b and c.

^b See Table 1, note a.

^c As part of their debriefing, participants were asked the following: "Please rate how important each of the following perspectives on the balanced scorecard was in your assessment of the success of the strategy at Paladin Pizza by allocating 100 points among the four perspectives, allocating more points to perspectives that were more important in your assessment." The names of the four scorecard perspectives were then provided with a blank next to each perspective for participants' point allocations.

^d The decrease in significance of the independent variable is significant (Freedman-Schatzkin t = 1.79, p = 0.040).

V. CONCLUSION

Experimental results show that the combination of involving managers in the selection of scorecard measures and framing the balanced scorecard as a causal chain can mitigate optimistic assessments of strategies. When managers use balanced-scorecard data to evaluate the success of a strategic initiative that they were involved in selecting, they perceive the initiative as more successful than managers who were not involved in the initiative-selection process. However, managers who are involved in measure selection appear less affected by their involvement in initiative selection when they are provided with the predicted causal chain of performance. These results highlight the importance of a causal-chain framing of the scorecard, which balanced-scorecard proponents emphasize but practitioners often fail to implement, as well as potential benefits of manager involvement in measure selection.

I find no evidence that simply framing the balanced scorecard as a causal chain is sufficient to overcome managers' motivated reasoning in a strategy-evaluation task. This non-effect suggests that, even in this simple setting, merely framing the scorecard as a causal-chain does not provide enough guidance to overcome motivated reasoning. It seems unlikely that it would do so in more realistic settings, in which greater complexity would allow managers even more flexibility to interpret results favorably. However, involving managers in the selection of performance measures provides a strong countervailing form of motivated reasoning that, when managers have a causal-chain focus, is able to improve strategy evaluation. Future research could examine whether the effects of framing the scorecard as a causal chain (absent involvement in measure selection) are stronger when managers are provided with additional causal-chain-related data (e.g., correlations), or when managers are given additional training on using the causal chain.

Results from my experiment suggest that motivated reasoning affects managers' emphasis on preference-inconsistent information, but information acquisition and information reliance (given information acquisition) could both drive these results. Future research could more closely examine the process through which managers' motivated reasoning works, examining whether involvement in measure selection coupled with a causal-chain framing increases managers' reliance on preference-inconsistent information they encounter or their acquisition of that information. In general, arriving at a better understanding of the motivated reasoning process could potentially guide methods such as decision aids used to mitigate the effects of motivated reasoning in specific settings.

These results speak to the use of the balanced scorecard as a tool for engendering a longer-term focus in managers. In their early work on the subject, [Kaplan and Norton \(1996a, 1996b\)](#) suggest that emphasizing leading measures, in conjunction with lagging measures, could curtail some of the managerial myopia seen in practice, where all that counts is *today's* bottom line. My results indicate that manager involvement in initiative selection will lead to more emphasis on the direct effects these initiatives have on leading performance indicators. However, the results also indicate that care is needed (through manager involvement in measure selection and framing the scorecard as a causal chain) to ensure that motivated reasoning does not lead to another form of myopia, where only preference-consistent evidence is examined in performance analyses.

The findings of this study relate to prior work on the "escalation of commitment," the tendency for individuals to increase commitment to a failing course of action (see [Staw \[1981\]](#) for a review). While motivated reasoning processes can result in an escalation of commitment, escalation of commitment does not necessarily imply that motivated reasoning has taken place, as escalation of commitment does not speak to the data-evaluation process. An escalation-of-commitment outcome could result from a conscious desire to save face, investing more in the hopes of recovering losses, in addition to being driven by motivated reasoning. Though the primary results of the current study are consistent with an escalation of commitment to an unsuccessful strategic initiative, process data and mediation results suggest that motivated reasoning is

a driving force behind this escalation. Future research could further disentangle motivated reasoning from more conscious processes that lead to an escalation of commitment.

This study has multiple limitations. First, I investigate only two types of scorecard implementation involvement in investigating the effects of motivated reasoning on strategy-evaluation judgments. Manager involvement in the assignment of performance targets, the selection of strategic objectives, or the development of the hypothesized causal chain itself will also likely influence these judgments. Second, my experiment held constant that scorecard feedback indicated a strong direct performance link and no indirect link. Alternate feedback realizations could potentially lead to different interactions with motivated reasoning processes. Third, participants in my study make judgments individually. Thus, results may not generalize to settings where strategy evaluation is done in groups. Evidence from psychology suggests that results may intensify with group decision making (e.g., [Schulz-Hardt et al. 2000](#)), though different levels of involvement in initiative and measure selection among group members could mitigate these effects ([Schulz-Hardt et al. 2006](#)). Fourth, participants in my study were students with little experience using the balanced scorecard. Though this feature of the subject pool allows for a more powerful manipulation of participants' understanding of the causal chain, it potentially calls into question inferences relating to more experienced scorecard users. However, research in psychology suggests that even experts fall prey to the effects of motivated reasoning ([Cuccia et al. 1995](#); [Hackenbrack and Nelson 1996](#); [Cloyd and Spilker 1999](#); [Phillips 2002](#); [Beeler and Hunton 2002](#); [Wilks 2002](#); [Kadous et al. 2003](#); [Han et al. 2006](#); [Moore et al. 2006](#); [Nelson 2006](#); [Kadous et al. 2008](#)).

This study contributes to accounting research by providing important insights into the balanced-scorecard implementation process. Prior empirical research on the balanced scorecard has focused primarily on scorecard usage, holding constant issues relating to scorecard implementation. This study explicitly manipulates manager involvement in the scorecard implementation process, and shows how motivated reasoning impacts evaluations when managers are involved in the selection of scorecard initiatives. Further, results suggest that a causal view of the scorecard, in conjunction with involvement in scorecard measure selection, helps mitigate these effects. Finally, unlike the majority of empirical research on the balanced scorecard, this study focuses on the use of the scorecard for developing and evaluating strategy, currently a major emphasis among scorecard proponents (e.g., [Kaplan and Norton 2001](#)).

REFERENCES

- American Quality Foundation. 1992. *Best Practices Report*. Cleveland, OH: Ernst & Young.
- Anderson, E. W., C. Fornell, and D. R. Lehmann. 1994. Customer satisfaction, market share, and profitability—Findings from Sweden. *Journal of Marketing* 58 (3): 53–66.
- Atkinson, A. A., R. Balakrishnan, P. Booth, J. M. Cote, T. Groot, T. Malmi, H. Roberts, E. Uliana, and A. Wu. 1997. New directions in management accounting research. *Journal of Management Accounting Research* 9: 79–108.
- Banker, R. D., H. Chang, and M. J. Pizzini. 2004. The balanced scorecard: Judgmental effects of performance measures linked to strategy. *The Accounting Review* 79 (1): 1–23.
- , G. Potter, and D. Srinivasan. 2000. An empirical investigation of an incentive plan that includes nonfinancial performance measures. *The Accounting Review* 75 (1): 65–92.
- Baron, R. M., and D. A. Kenny. 1986. The moderator mediator variable distinction in social-psychological research—Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51 (6): 1173–1182.
- Bazerman, M. H., G. Loewenstein, and D. A. Moore. 2002. Why good accountants do bad audits. *Harvard Business Review* 80 (11): 96–102.
- , K. P. Morgan, and G. F. Loewenstein. 1997. The impossibility of auditor independence. *Sloan Management Review* 38 (4): 89–94.

- Beeler, J. D., and J. E. Hunton. 2002. Contingent economic rents: Insidious threats to audit independence. *Advances in Accounting Behavioral Research* 5: 21–50.
- Bloomfield, R. J., and J. L. Luft. 2006. Responsibility for cost management hinders learning to avoid the winner's curse. *The Accounting Review* 81 (1): 29–47.
- Boiney, L. G., J. Kennedy, and P. Nye. 1997. Instrumental bias in motivated reasoning: More when more is needed. *Organizational Behavior and Human Decision Processes* 72 (1): 1–24.
- Bonner, S. E. 2007. *Judgment and Decision Making in Accounting*. Upper Saddle River, NJ: Prentice Hall.
- , R. Libby, and M. W. Nelson. 1996. Using decision aids to improve auditors' conditional probability judgments. *The Accounting Review* 71 (2): 221–240.
- , ———, and ———. 1997. Audit category knowledge as a precondition to learning from experience. *Accounting, Organizations and Society* 22 (5): 387–410.
- Butt, J. L. 1988. Frequency judgments in an auditing-related task. *Journal of Accounting Research* 26 (2): 315–330.
- Buytendijk, F., B. Wood, and L. Geishecker. 2004. Drivers and challenges of corporate performance measurement. No. R-22-0730. Stamford, CT: Gartner, Inc.
- Campbell, D., S. M. Datar, S. L. Kulp, and V. G. Narayanan. 2008. Testing strategy with multiple performance measures: Evidence from a balanced scorecard at Store24. Working paper, Harvard Business School.
- Cloyd, C. B., and B. C. Spilker. 1999. The influence of client preferences on tax professionals' search for judicial precedents, subsequent judgments and recommendations. *The Accounting Review* 74 (3): 299–322.
- Cokins, G. 2005. Performance management: Making it work: The promise and perils of the balanced scorecard. *DM Review Online* (August 2005).
- Cuccia, A. D., K. Hackenbrack, and M. W. Nelson. 1995. The ability of professional standards to mitigate aggressive reporting. *The Accounting Review* 70 (2): 227–248.
- Dawson, E., T. Gilovich, and D. T. Regan. 2002. Motivated reasoning and performance on the Wason selection task. *Personality and Social Psychology Bulletin* 28 (10): 1379–1387.
- , ———, and ———. 2006a. Motivated reasoning and susceptibility to the “cell A” bias. Working paper, Yale University and Cornell University.
- , K. Savitsky, and D. Dunning. 2006b. “Don't tell me, I don't want to know:” Understanding people's reluctance to obtain medical diagnostic information. *Journal of Applied Social Psychology* 36 (3): 751–768.
- Dilla, W. N., and P. J. Steinbart. 2005. Relative weighting of common and unique balanced scorecard measures by knowledgeable decision makers. *Behavioral Research in Accounting* 17: 43–53.
- Ditto, P. H., and D. F. Lopez. 1992. Motivated skepticism—Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology* 63 (4): 568–584.
- , G. D. Munro, A. M. Apanovitch, J. A. Scepansky, and L. K. Lockhart. 2003. Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. *Personality and Social Psychology Bulletin* 29 (9): 1120–1132.
- , J. A. Scepansky, G. D. Munro, A. M. Apanovitch, and L. K. Lockhart. 1998. Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology* 75 (1): 53–69.
- Dunning, D., J. A. Meyerowitz, and A. D. Holzberg. 1989. Ambiguity and self-evaluation—The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology* 57 (6): 1082–1090.
- Farrell, A. M., K. Kadous, and K. L. Towry. 2009. Melioration and perfectionism: Effects of information about internalities on employee effort allocations in multi-period tasks. Working paper, The University of Illinois at Urbana-Champaign and Emory University.
- Freedman, L. S., and A. Schatzkin. 1992. Sample size for studying intermediate endpoints within intervention trials or observational studies. *American Journal of Epidemiology* 136 (9): 1148–1159.
- Frigo, M. L., and K. R. Krumwiede. 2000. The balanced scorecard. *Strategic Finance* 81 (7): 50–54.
- Gilovich, T. 1991. *How we know what isn't so: The fallibility of human reason in everyday life*. New York, N.Y.: Free Press.

- Hackenbrack, K., and M. W. Nelson. 1996. Auditors' incentives and their application of financial accounting standards. *The Accounting Review* 71 (1): 43–59.
- Hales, J. 2007. Directional preferences, information processing, and investors' forecasts of earnings. *Journal of Accounting Research* 45 (3): 607–628.
- , D. V. Moser, and J. G. Birnberg. 2006. Revisiting how effectively the audit review process reduces judgment biases. Working paper, Rutgers University at Camden and University of Pittsburgh.
- Herrnstein, R. J. 1990. Rational choice theory—Necessary but not sufficient. *The American Psychologist* 45 (3): 356–367.
- , and D. Prelec. 1991. Melioration—A theory of distributed choice. *The Journal of Economic Perspectives* 5 (3): 137–156.
- , and ———. 1992. Melioration. In *Choice Over Time*, edited by Loewenstein, G., and J. Elster, 235–263. New York, NY: Russell Sage Foundation.
- Hunton, J. E., and R. A. McEwen. 1997. An assessment of the relation between analysts' earnings forecast accuracy, motivational incentives and cognitive information search strategy. *The Accounting Review* 72 (4): 497–515.
- Ittner, C. D., R. A. Lambert, and D. F. Larcker. 2003a. The structure and performance consequences of equity grants to employees of new economy firms. *Journal of Accounting and Economics* 34 (1–3): 89–127.
- , and D. F. Larcker. 2003. Coming up short on nonfinancial performance measurement. *Harvard Business Review* 81 (11): 88–95.
- , ———, and M. W. Meyer. 2003b. Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review* 78 (3): 725–758.
- James, L. R., and J. M. Brett. 1984. Mediators, moderators, and tests for mediation. *The Journal of Applied Psychology* 69 (2): 307–321.
- , S. A. Mulaik, and J. M. Brett. 2006. A tale of two methods. *Organizational Research Methods* 9 (2): 233–244.
- Kadous, K., S. J. Kennedy, and M. E. Peecher. 2003. The effect of quality assessment and directional goal commitment on auditors' acceptance of client-preferred accounting methods. *The Accounting Review* 78 (3): 759–778.
- , A. M. Magro, and B. C. Spilker. 2008. Do effects of client preference on accounting professionals' information search and subsequent judgments persist with high practice risk? *The Accounting Review* 83: 133–156.
- Kaplan, R. S., and D. P. Norton. 1992. The balanced scorecard—Measures that drive performance. *Harvard Business Review* 70 (1): 71–79.
- , and ———. 1996a. *The Balanced Scorecard: Translating Strategy into Action*. Boston, MA: Harvard Business School Press.
- , and ———. 1996b. Using the balanced scorecard as a strategic management system. *Harvard Business Review* 74 (1): 75–85.
- , and ———. 2000. Having trouble with your strategy? Then map it. *Harvard Business Review* 78 (5): 167–176.
- , and ———. 2001. *The Strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment*. Boston, MA: Harvard Business School Press.
- , and ———. 2004a. Measuring the strategic readiness of intangible assets. *Harvard Business Review* 82 (2): 52–63.
- , and ———. 2004b. *Strategy Maps: Converting Intangible Assets into Tangible Outcomes*. Boston, MA: Harvard Business School Press.
- , and ———. 2006. *How to Implement a New Strategy Without Disrupting Your Organization*. Boston, MA: Harvard Business Review.
- Krishnan, R., J. Luft, and M. D. Shields. 2005. Effects accounting-method choices on subjective performance-measure weighting decisions: Experimental evidence on precision and error covariance. *The Accounting Review* 80 (4): 1163–1192.
- Kunda, Z. 1990. The case for motivated reasoning. *Psychological Bulletin* 108 (3): 480–498.
- Libby, R., R. Bloomfield, and M. W. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27 (8): 777–812.

- Libby, T., S. E. Salterio, and A. Webb. 2004. The balanced scorecard: The effects of assurance and process accountability on managerial judgment. *The Accounting Review* 79 (4): 1075–1094.
- Lipe, M. G., and S. E. Salterio. 2000. The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review* 75 (3): 283–298.
- , and ———. 2002. A note on the judgmental effects of the balanced scorecard's information organization. *Accounting, Organizations and Society* 27 (6): 531–540.
- Lord, C. G., L. Ross, and M. R. Lepper. 1979. Biased assimilation and attitude polarization—Effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37 (11): 2098–2109.
- MacKinnon, D. P., C. M. Lockwood, J. M. Hoffman, S. G. West, and V. Sheets. 2002. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods* 7 (1): 83–104.
- Mainwaring, L. 1997. Maximisation and melioration as alternative forms of firm behavior. *Journal of Economic Behavior & Organization* 32 (3): 395–411.
- Malina, M. A., and F. H. Selto. 2001. Communicating and controlling strategy: An empirical study of the effectiveness of the balanced scorecard. *Journal of Management Accounting Research* 13: 47.
- , and ———. 2004. Causality in a performance measurement model. Working paper, University of Colorado at Denver and University of Colorado at Boulder.
- Moore, D. A., P. E. Tetlock, L. Tanlu, and M. H. Bazerman. 2006. Conflicts of interest and the case of auditor independence: Moral seduction and strategic issue cycling. *Academy of Management Review* 31 (1): 10–29.
- Morecroft, J. D. W., R. Sanchez, and A. Heene. 2002. *Systems Perspectives on Resources, Capabilities, and Management Processes*. 1st edition. Amsterdam, NY: Pergamon.
- Nelson, M. W. 1993. The effects of error frequency and accounting knowledge on error diagnosis in analytical review. *The Accounting Review* 68 (4): 804–824.
- . 2006. Ameliorating conflicts of interest in auditing: Effects of recent reforms on auditors and their clients. *Academy of Management Review* 31 (1): 30–42.
- , R. Libby, and S. E. Bonner. 1995. Knowledge structure and the estimation of conditional probabilities in audit planning. *The Accounting Review* 70 (1): 27–47.
- , and W. B. Tayler. 2007. Information pursuit in financial statement analysis: Effects of choice, effort, and reconciliation. *The Accounting Review* 82 (3): 731–758.
- Niven, P. R. 2002. *Balanced Scorecard Step by Step: Maximizing Performance and Maintaining Results*. New York, NY: Wiley.
- Phillips, F. 2002. The distortion of criteria after decision-making. *Organizational Behavior and Human Decision Processes* 88 (2): 769–784.
- Pyszczynski, T., and J. Greenberg. 1987. Toward an integration of cognitive and motivational perspectives on social inference—A biased hypothesis-testing model. *Advances in Experimental Social Psychology* 20: 297–340.
- Rigby, D., and B. Bilodeau. 2009. *Management Tools and Trends 2009*. Boston, MA: Bain & Company.
- Roberts, M. L., T. L. Albright, and A. R. Hibbets. 2004. Debiasing balanced scorecard evaluations. *Behavioral Research in Accounting* 16: 75–88.
- Schulz-Hardt, S., D. Frey, C. Luthgens, and S. Moscovici. 2000. Biased information search in group decision making. *Journal of Personality and Social Psychology* 78 (4): 655–669.
- , F. C. Brodbeck, A. Mojzisch, R. Kerschreiter, and D. Frey. 2006. Group decision making in hidden profile situations: Dissent as a facilitator for decision quality. *Journal of Personality and Social Psychology* 91 (6): 1080–1093.
- Smith, R. E., and W. F. Wright. 2004. Determinants of customer loyalty and financial performance. *Journal of Management Accounting Research* 16: 183–205.
- Seybert, N., and R. J. Bloomfield. 2009. Contagion of wishful thinking in markets. *Management Science* 55 (5): 738–751.
- Staw, B. M. 1981. The escalation of commitment to a course of action. *Academy of Management Review* 6 (4): 577–587.

- Thayer, J. 2009. Value relevance or valence? The effect of shareholders' preferences on their information-seeking behavior. Working paper, The University of Georgia.
- Vera-Muñoz, S. C., M. Shackell-Dowell, and M. Buehner. 2007. Accountants' usage of causal business models in the presence of benchmark data: A note. *Contemporary Accounting Research* 24: 1015–1038.
- Wilks, T. J. 2002. Predecisional distortion of evidence as a consequence of real-time audit review. *The Accounting Review* 77 (1): 51–71.

Copyright of Accounting Review is the property of American Accounting Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.